

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/29278>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

ART 2-A for Optimal Test Series Design in QSAR[§]

Daniel Domine,^{*,†} James Devillers,[†] Dietrich Wienke,[‡] and Lutgarde Buydens[‡]

CTIS, 21 rue de la Bannière, 69003 Lyon, France, and Catholic University of Nijmegen,
Laboratory for Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Received April 26, 1996[®]

The family of adaptive resonance theory (ART) based systems concerns distinct artificial neural networks for unsupervised and supervised clustering analysis. Among them, the ART 2-A paradigm presents numerous strengths for data analysis. After a rapid presentation of the ART 2-A theory and algorithmic information, the usefulness of this neural network for the selection of optimal test series is estimated. The results are compared with those obtained from hierarchical cluster analysis and visual mapping methods. The advantages and drawbacks of each method are discussed. We show that ART 2-A represents a new useful nonlinear statistical tool for QSAR and drug design.

1. INTRODUCTION

The design of valuable structure–activity or structure–property relationships requires the use of training and testing sets with compounds presenting a maximal “meaningful variety”. Even if many techniques for selecting sets of compounds presenting a high information content with respect to the activity or property studied are available, these approaches are not completely satisfactory since they present various drawbacks. Indeed, for example, some can be considered as black boxes since they only provide optimal sets without any information on the similarities with the other compounds available.¹ In the same way, the research for new biologically active chemicals in agrochemistry and medicinal chemistry or decision making in the management of toxic substances requires inspection of several thousands of chemicals for which only limited information is available. Since it is practically impossible to test all the possible candidates, a decision must be taken as where to start from. This implies the adoption of strategies for a rational selection of the most relevant compounds for preliminary biological testing and for the derivation of (quantitative) structure–activity relationships ((Q)SAR) and quantitative structure–property relationships (QSPR). This kind of strategy is known as optimal test series design. A plethora of publications dealing with this crucial problem can be found in the literature (see for example the review paper of Pleiss and Unger²). Significant advances in the field dealt with the use of clustering and display linear multivariate methods allowing selection of test series by simple visual inspection of 2-D plots summarizing the information content of a matrix of physicochemical properties.^{3–9} More recently, nonlinear approaches such as nonlinear mapping (NLM),^{1,10,11} nonlinear neural mapping (N2M),¹² and genetic algorithms¹³ have been proposed opening new perspectives in the field.

The usefulness of adaptive resonance theory (ART) based neural networks in chemometrics has recently been stressed.^{14–22} Among them, ART 2-A, an algorithmic

version of ART 2, has been shown to present some advantages over classical clustering tools such as rapid training speed, self-organization behavior, and interpretability of the network weights. Under these conditions, the aim of this paper is to explore the interest of ART 2-A as a new alternative for solving the complex problem of optimal test series selection not only due to its ability to perform clustering but also due to the fact that it provides useful information on the clusters formed and their relationships.

2. THEORY OF ART 2-A

By means of his adaptive resonance theory, Grossberg^{23,24} started to model selected aspects of real-time pattern recognition in the brain. This theory rather provided neural models that were not explicitly designed for multivariate data analysis. ART can explain how a pattern is classified by a trained biological neural network, when this pattern does not fit into any existing learned category. ART states for this typical extrapolation case that a brain extends its knowledge by an initialization of up to here unused regions of “fresh” neurons. The new region is dynamically linked to the active region of the brain. In the brain, this ability avoids collapses when confronted, for example, with an unexpected visual novelty. This novelty is rather dynamically detected as deviating from all recorded categories of knowledge, and it is immediately (in real-time) stored by the brain as a new knowledge. The research group of Grossberg and Carpenter then developed several ART based classifiers. The methods mainly differ in approaching pattern similarity by either Euclidean angle based or fuzzy theory based distance concepts. Some ART classifiers such as ART 1^{23–25} and ARTMAP²⁶ are restricted to binary input vectors while ART 2,²⁷ ART 2-A,²⁸ ART 3,²⁹ and FuzzyART³⁰ can stably learn to categorize either binary or analog input patterns. ART 3 can additionally carry out parallel search, or hypothesis testing, of distributed recognition codes in a multilevel network hierarchy. FuzzyARTMAP³¹ can rapidly learn stable categorical mappings between analog or binary input and output vectors. Historically, the group of unsupervised ART methods ART 1,^{23–25} ART 2,²⁷ ART 3,²⁹ and FuzzyART³⁰ was developed before the supervised ART techniques ARTMAP²⁶ and FuzzyARTMAP.³¹ Many of the basic papers on the topic have been compiled in a seminal book

* Author to whom all correspondence should be addressed.

† CTIS.

‡ Catholic University of Nijmegen.

[§] Key words: adaptive resonance theory; ART networks; hierarchical cluster analysis; nonlinear mapping; nonlinear neural mapping; selection of test series.

[®] Abstract published in *Advance ACS Abstracts*, November 15, 1996.

by Carpenter and Grossberg³² dedicated to self-organizing neural networks. In the present case study, we used ART 2-A, a simplified ART 2 algorithm, due to its ability to deal with analog input patterns, its ease for computation, and also its learning speed.²⁸ Like all ART paradigms, the mathematical principles behind ART 2-A are rather complicated but the fundamental ideas are not. Under these conditions, we only outline below the main concepts of this paradigm and provide algorithmic information. Readers that are more mathematically inclined will find an abundance of theory and mathematical demonstrations in the original articles of Carpenter and co-workers.²³⁻³² The unsupervised clustering performed by ART 2-A can be briefly summarized as follows. ART 2-A is divided into two subsystems: an attentional subsystem, which processes familiar stimuli, and an orienting subsystem that detects unfamiliar input patterns and resets the attentional subsystem when it detects such a pattern. It compares input patterns with the existing network structure to check whether they fit into ("resonate with") it. In case of resonance, an "adaptation" of the existing weights is performed, otherwise the structure is adapted by creation of new weight vector(s). Hence, the network learns by modification or creation of a set of coefficients (weights). Thus, ART 2-A like the other ART networks is designed to resolve the stability-plasticity dilemma. It is stable enough to preserve past learning, but nevertheless remains adaptable enough to incorporate new information whenever it might appear.

In practice, an input pattern presented to the network is first preprocessed and then compared to each of the existing cluster prototypes. Among them, the winner is the prototype that is the most similar to the input pattern. If this similarity exceeds a predefined limit termed vigilance parameter, then the network learns the input by modifying the winner's weights to more closely reflect the input. If the similarity is less than required by the vigilance parameter, the network adapts its structure by adding a new neuron whose weights become identical to the input pattern presented. At the end of the learning process, the network has classified the input patterns into a variable number of clusters depending on the value of the vigilance parameter and on the structure of the data set at hand.

The algorithm used to perform the above clustering and the structure of an ART 2-A network are depicted in Figures 1 and 2, respectively. The different steps of the algorithm are detailed below. Applied in a QSAR context, ART 2-A allows one to derive clusters from a data matrix \mathbf{X} of n individuals described by m variables. Each individual \mathbf{x}_i is therefore considered as a vector in the m -dimensional space of variables. Some of the individuals can be closer to each other in this space forming groups if the values of their m variables are similar. The goal is to find such groups (clusters) of similar individuals, whereby the number c of groups in the data set \mathbf{X} is not known *a priori*. The ART 2-A algorithm incorporates the basic features of all ART systems, notably, pattern matching between bottom-up input and top-down learned prototype vectors. There are bottom-up and top-down connections between each input and output neurons (layers F1 and F2, respectively) which comprise the "adaptive filters". In addition, a preprocessing layer (F0) is added. It performs a normalization and allows contrast enhancement between the input patterns (Figure 2).

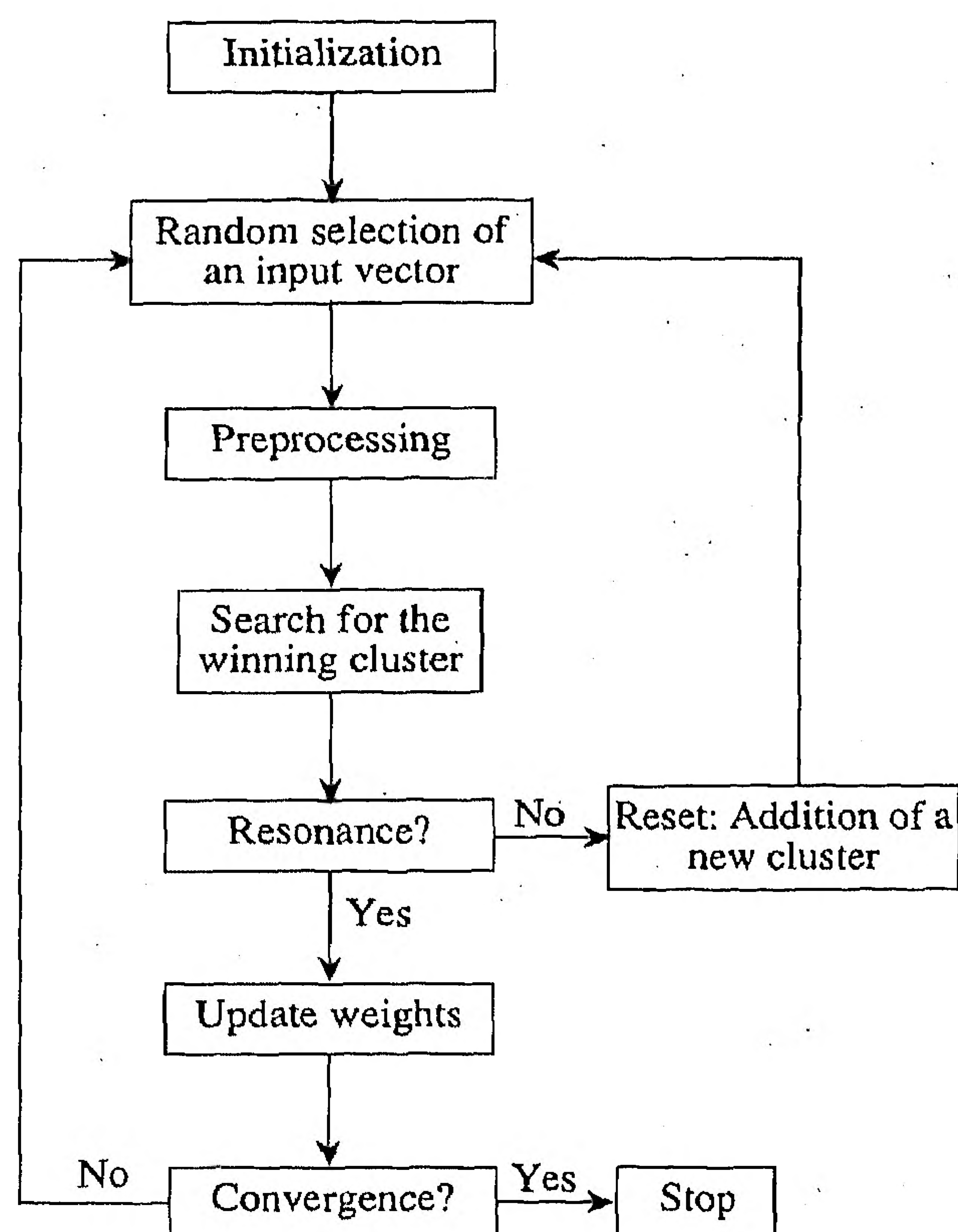


Figure 1. Algorithm flow diagram of ART 2-A.

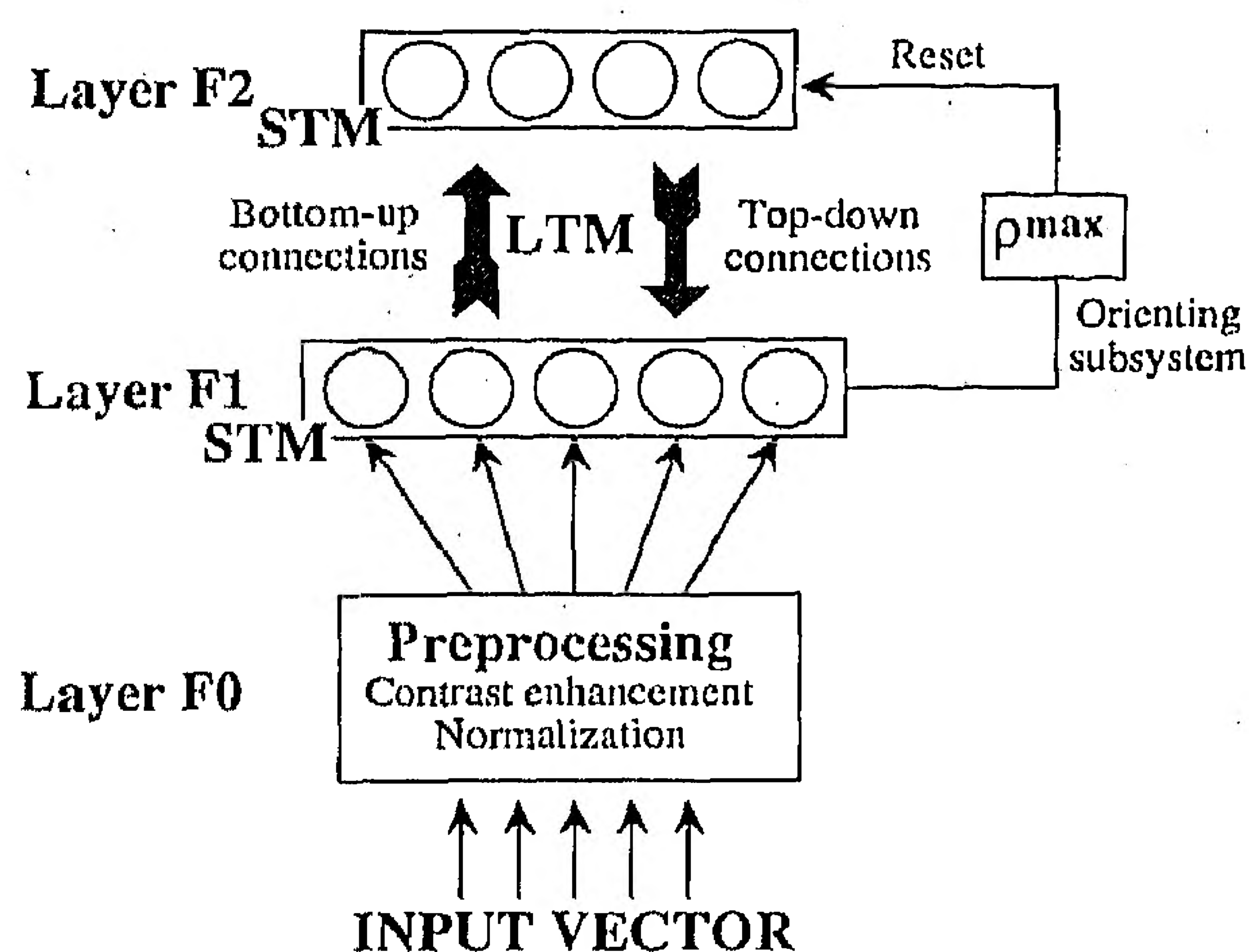


Figure 2. Architecture of an ART 2-A network.

An initialization of the so-called long-term memory (LTM) which is the matrix of weight coefficients is performed. The bottom-up weights are assigned a same value $1/\sqrt{m}$, while top-down weights are initially set to 0. The number of potential categories is arbitrary, and initially each category is said to be "uncommitted". The different learning parameters discussed below (i.e., α , θ , ρ^{\max} , η) are also set during this initialization phase. Then, a single individual \mathbf{x}_i from the data matrix \mathbf{X} is randomly selected and preprocessed to yield \mathbf{r}_i . This consists in a normalization of the individual to unit length (i.e., the norm of each individual \mathbf{x}_i equals one) and a contrast enhancement. Normalization and contrast enhancement are performed by means of eqs 1 and 2, respectively.

$$\mathbf{o}_i = \mathbf{x}_i / \|\mathbf{x}_i\| \quad (1)$$

$$F_{\theta}(\mathbf{o}_i) \equiv \begin{cases} r_{ij} = o_{ij} & \text{if } o_{ij} > \theta \\ r_{ij} = 0 & \text{otherwise} \end{cases} \quad (2)$$

where θ is a threshold that smoothes out low values and satisfies the following inequality $0 < \theta \leq 1/\sqrt{m}$.

The F1 layer then takes the F0 output and gates it to the F2 layer through the bottom-up connections (\mathbf{z}_k). During this step, the input (short-term memory (STM)) is compared with all the already existing $k = 1$ to c clusters (i.e., committed nodes) and the uncommitted nodes of the F2 layer. Indeed, each cluster on the F2 layer is represented by a weight vector \mathbf{z}_k^* having the same dimension as the input vectors. The similarity between the input vectors and the F2 nodes is expressed as shown in eq 3.

$$\rho_k^{\text{calc}} = \begin{cases} \alpha \sum_j r_{ij} & \text{if } k \text{ is an uncommitted node} \\ \mathbf{r}_i \mathbf{z}_k^* & \text{if } k \text{ is a committed node} \end{cases} \quad (3)$$

The constant α satisfies $\alpha \leq 1/\sqrt{m}$. Because \mathbf{r}_i and \mathbf{z}_k^* have unit length, eq 3 comes down to the cosine of the angle between both vectors.

A winning neuron is selected among the F2 nodes. It corresponds to this presenting the maximal similarity (ρ_k^{calc}). In the original algorithm, if more than one node is maximum, then one chooses one at random. In our study, we chose this having the lowest index k . The selected node is said to be committed. In the neural network terminology, the search for the winning neuron followed by its training is called competitive learning. This notion is also known, for example, in Kohonen networks.^{33,34} However, unlike Kohonen networks, in ART 2-A only the weights of the winning node can be updated. In addition, the winning neuron is not updated if its ρ_k^{calc} value (noted $\rho^{\text{calc}, \text{winner}}$) is inferior to the vigilance parameter noted ρ^{max} (graphically expressed, ρ^{max} forms the spatial limit of each cluster). Indeed, the next step is the resonance check. During this step, $\rho^{\text{calc}, \text{winner}}$ is compared with ρ^{max} where ρ^{max} is constrained so that $0 \leq \rho^{\text{max}} \leq 1$. If k is committed and $\rho^{\text{calc}, \text{winner}} < \rho^{\text{max}}$, then k is reset to the index of an uncommitted node. A large ρ^{max} will thus provide many new clusters having a small diameter. In the reverse case where ρ^{max} is chosen close to 0, only a few but large clusters are generated. If the selected neuron is an uncommitted node as it is necessarily the case at the very beginning of an ART 2-A run, when the first individual is randomly selected and presented to the network, then \mathbf{r}_i is simply copied to \mathbf{z}_k^* . Otherwise, the network weights are updated as shown below:

$$\mathbf{z}_k^*{}^{\text{new}, \text{winner}} = \mathbf{t}_k / \|\mathbf{t}_k\| \quad (4)$$

where

$$\mathbf{t}_k = \mathbf{u}_k + (1 - \eta) \mathbf{z}_k^*{}^{\text{old}, \text{winner}} \quad (5)$$

with

$$\mathbf{u}_k = \eta (\mathbf{v}_k / \|\mathbf{v}_k\|) \quad (6)$$

and

$$v_{kj} = r_{ij} \quad \text{if } z_{kj}^*{}^{\text{old}, \text{winner}} > \theta \quad (7)$$

$$v_k = 0 \text{ otherwise}$$

θ is a threshold that smoothes out low values providing in this way a nonlinear transfer of the weight vector. The aim of this update is to make the spatial position of the winning

cluster closer to the presented \mathbf{r}_i . In practice, η is chosen between 0 and 1. An η value close to 0 provides small stepwise changes in the weights. Thus, via "weight adaptation", an ART network stores a weighted part of the present input individual in the LTM. If the resonance condition (i.e., $\rho_k^{\text{calc}} > \rho^{\text{max}}$) is not satisfied, the network adapts in place of its weights first its structure toward the discovered novelty. "Structure adaptation" means adding a new cluster (additional weight vector or neuron \mathbf{z}_{c+1}^*) to the existing ones. The significantly deviating novel individual \mathbf{r}_i is stored immediately in this additional neuron by direct copy. This is another original idea of Grossberg: ART neural networks use not only their weights but also their structure for information storage.

After this step of adaptation, another input vector is randomly selected from the training set, and the entire process of "resonance" and "adaptation" is repeated, whereby the content of the STM is continuously overwritten by the new input vectors. When the random sampling of the training data matrix \mathbf{X} has been performed n times, one has performed one epoch. Simultaneously, the contrast between clusters in the LTM increases. The process converges within a few epochs with formation of c clusters. The chosen constant of vigilance parameter ρ^{max} determines how many clusters c are formed. In other words, by a suitable choice of ρ^{max} , the number c of clusters to derive from the multivariate data cloud \mathbf{X} can be monitored. In this way, a variation of ρ^{max} can serve for active exploration of a data set \mathbf{X} by resolving it into distinct numbers of clusters presenting various sizes.

With ART, the more compact and well separated the clusters are, the less subjective the user's choice of η and ρ is and the more stable the results are. In this situation, ART will always find the "true" number c of hidden clusters. At the opposite, the more the data scatter or the more they form a continuous hypersurface of equidistant points, the higher the influence of η and ρ on the clustering result is.

Only a few papers dealing with applications of ART 2-A can be found in the literature, but they can provide complementary information on the ART 2-A paradigm and can represent a source of ideas for the development of ART 2-A applications.^{16,18-22,35-39}

3. EXPERIMENTAL SECTION

ART 2-A was applied on a data matrix of 103 aliphatic substituents (Table 1) described by the hydrophobic constant for aliphatic substituents Fr, H-bonding acceptor (HBA) and donor (HBD) abilities, the molar refractivity (MR), and the inductive parameter F, respectively.⁴⁰ The selection of this data set was directed by the fact that the results could be compared with a series of methods so that the advantages and drawbacks of ART 2-A could be more easily evidenced. Indeed, this matrix was first analyzed by means of hierarchical cluster analysis (HCA).⁴⁰ In addition, we recently proposed the use of NLM for obtaining a single easily interpretable map summarizing the information contained in the original data table.¹⁰ More recently the use of a hybrid system combining clustering and display techniques was also experienced.¹² The data were scaled by normalization in the range [0, 1] (i.e., using a min/max equation) per column. A quantitative analysis of the network training was performed by running the network with different values of the vigilance parameter and by monitoring the number of clusters formed

Table 1. Aliphatic Substituents

| no. | substituent | no. | substituent | no. | substituent |
|-----|----------------------------------|-----|---|-----|---|
| 1 | Br | 36 | OCOCH ₃ | 70 | COC ₃ H ₇ |
| 2 | Cl | 37 | CO ₂ CH ₃ | 71 | OCOC ₃ H ₇ |
| 3 | F | 38 | NHCOCH ₃ | 72 | CO ₂ C ₃ H ₇ |
| 4 | I | 39 | C=O(NHCH ₃) | 73 | (CH ₂) ₃ CO ₂ H |
| 5 | NO ₂ | 40 | CH ₂ CH ₃ | 74 | NHCOC ₃ H ₇ |
| 6 | H | 41 | OCH ₂ CH ₃ | 75 | CONHC ₃ H ₇ |
| 7 | OH | 42 | CH ₂ OCH ₃ | 76 | C ₄ H ₉ |
| 8 | SH | 43 | SOC ₂ H ₅ | 77 | C(CH ₃) ₃ |
| 9 | NH ₂ | 44 | SC ₂ H ₅ | 78 | OC ₄ H ₉ |
| 10 | CBR ₃ | 45 | CH ₂ Si(CH ₃) ₃ | 79 | CH ₂ OC ₃ H ₇ |
| 11 | CCl ₃ | 46 | NHC ₂ H ₅ | 80 | NHC ₄ H ₉ |
| 12 | CF ₃ | 47 | N(CH ₃) ₂ | 81 | N(C ₂ H ₅) ₂ |
| 13 | CN | 48 | CH=CHCN | 82 | CH=CHCOC ₂ H ₅ |
| 14 | SCN | 49 | cyclopropyl | 83 | CH=CHCO ₂ C ₂ H ₅ |
| 15 | CO ₂ ⁻ | 50 | COC ₂ H ₅ | 84 | C ₅ H ₁₁ |
| 16 | CO ₂ H | 51 | CO ₂ C ₂ H ₅ | 85 | CH ₂ OC ₄ H ₉ |
| 17 | CH ₂ Br | 52 | OCOC ₂ H ₅ | 86 | C ₆ H ₅ |
| 18 | CH ₂ Cl | 53 | EtCO ₂ H | 87 | OC ₆ H ₅ |
| 19 | CH ₂ I | 54 | NHCO ₂ C ₂ H ₅ | 88 | SO ₂ C ₆ H ₅ |
| 20 | CONH ₂ | 55 | CONHC ₂ H ₅ | 89 | NHC ₆ H ₅ |
| 21 | CH=NOH | 56 | NHCOC ₂ H ₅ | 90 | 2-benzthiazolyl |
| 22 | CH ₃ | 57 | CH(CH ₃) ₂ | 91 | CH=CHCOC ₃ H ₇ |
| 23 | NHCONH ₂ | 58 | C ₃ H ₇ | 92 | CH=CHCO ₂ C ₃ H ₇ |
| 24 | OCH ₃ | 59 | OCH(CH ₃) ₂ | 93 | COC ₆ H ₅ |
| 25 | CH ₂ OH | 60 | OC ₃ H ₇ | 94 | CO ₂ C ₆ H ₅ |
| 26 | SOCH ₃ | 61 | CH ₂ OC ₂ H ₅ | 95 | OCOC ₆ H ₅ |
| 27 | OSO ₂ CH ₃ | 62 | SOC ₃ H ₇ | 96 | NHCOC ₆ H ₅ |
| 28 | SCH ₃ | 63 | SC ₃ H ₇ | 97 | CH ₂ C ₆ H ₅ |
| 29 | NHCH ₃ | 64 | NHC ₃ H ₇ | 98 | CH ₂ OC ₆ H ₅ |
| 30 | CF ₂ CF ₃ | 65 | Si(CH ₃) ₃ | 99 | CH ₂ Si(C ₂ H ₅) ₃ |
| 31 | C≡CH | 66 | 2-thienyl | 100 | CH=CHC ₆ H ₅ -(trans) |
| 32 | CH ₂ CN | 67 | 3-thienyl | 101 | CH=CHCOC ₆ H ₅ |
| 33 | CH=CHNO ₂ -(trans) | 68 | CH=CHCOCH ₃ | 102 | ferrocenyl |
| 34 | CH=CH ₂ | 69 | CH=CHCO ₂ CH ₃ | 103 | N(C ₆ H ₅) ₂ |
| 35 | COCH ₃ | | | | |

during learning in order to select an appropriate ρ^{\max} value. A full analysis of the ART 2-A results and comparison with previously published results using different techniques was performed by inspecting the numerical outputs and using graphical display techniques such as nonlinear mapping.⁴¹

4. RESULTS AND DISCUSSION

4.1. Setting the Cluster Size. Figure 3 illustrates for five different values of the vigilance parameter (ρ^{\max}) the number of clusters formed versus the number of cycles required for convergence with ART 2-A. It is noteworthy that in all cases convergence was obtained after approximately 10 epochs. This underlines the ability of ART networks to perform clustering very rapidly. This presents two main advantages. Indeed, the rapid convergence in training allows us to quickly optimize the network parameters, and, therefore, many configurations can be tested for exploratory data analysis. From Figure 3, it can also be deduced that the dominant clusters are always formed at the very beginning of the training. Indeed, the slope of the curves shown in Figure 3 are very high at the beginning, and the number of clusters formed then tends to a maximum very rapidly. Last, the low sensitivity between $\rho^{\max} = 0.1$ and $\rho^{\max} = 0.6$ indicates that there are basically three strong clusters in the structure of the data set. To compare our results with those previously published,^{10,12,40} the ρ^{\max} value was fixed so that 10 clusters could be obtained. This underlines again the advantage of this kind of network which allows a gradual inspection of a data set and also a monitoring of the precision required in the clustering. The appropriate ρ^{\max} value for the above constraint was found to

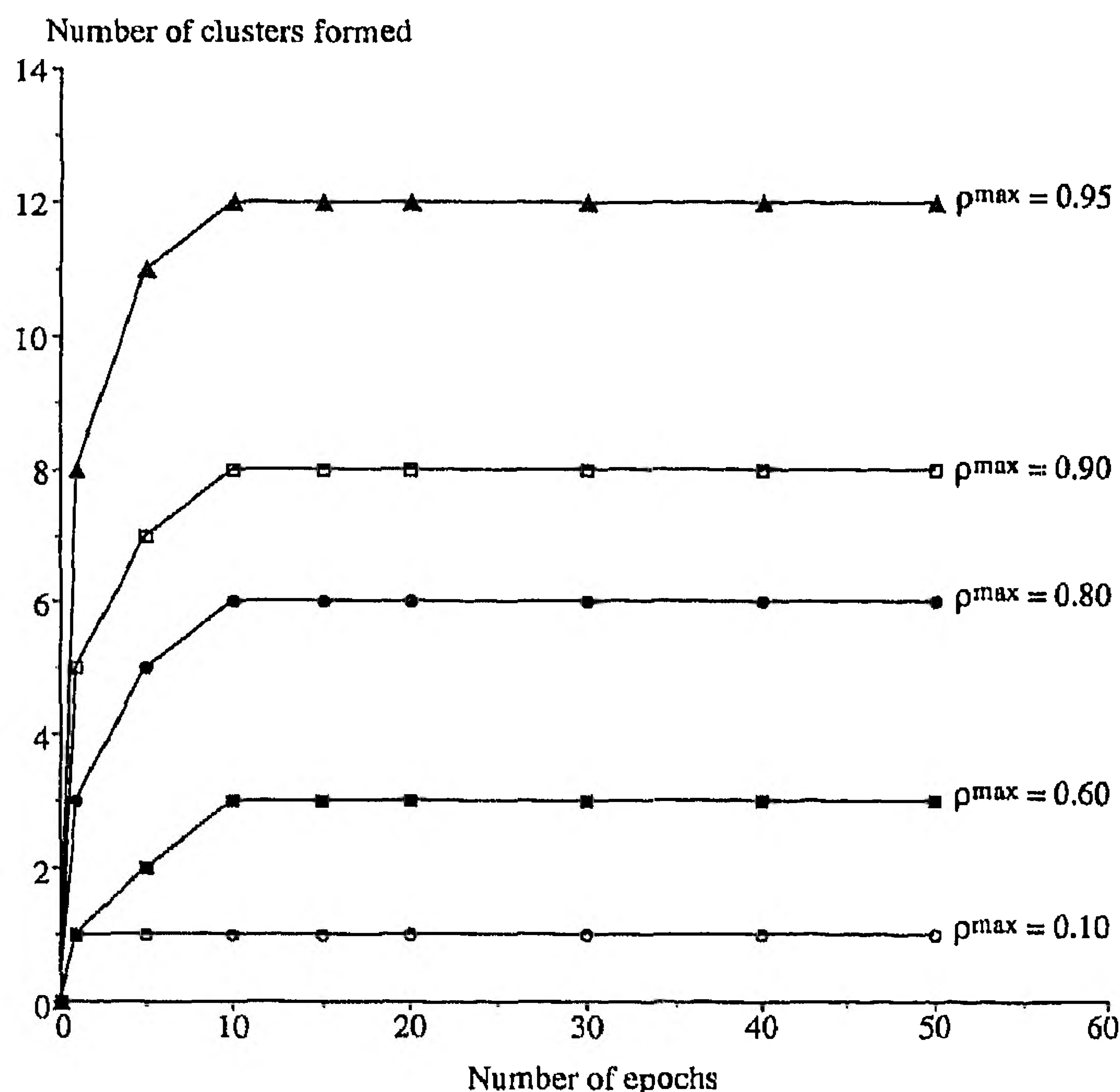


Figure 3. Plot of the number of clusters formed versus number of epochs for an ART 2-A neural network trained with the aliphatic data set at various ρ^{\max} values.

Table 2. Contents of the Ten ART 2-A Clusters^a

| cluster no. | substituents |
|-------------|--|
| 1 | 7, 9, 16, 20, 21, 23, 25, 38, 39, 54, 55, 56, 74, 75 |
| 2 | 29, 46, 53, 64, 73, 80, 89, 96 |
| 3 | 5, 13, 14, 24, 26, 27, 35, 36, 37, 43, 52, 62, 88 |
| 4 | 6, 10, 11, 17, 18, 19, 22, 28, 34, 40, 44, 49, 57, 58, 63, 66, |
| 5 | 32, 33, 41, 42, 47, 48, 50, 51, 59, 60, 61, 68, 69, 70, 71, 72, 78, 82, 87 |
| 6 | 15 |
| 7 | 45, 65, 67, 76, 77, 84, 86, 97, 99, 100, 102 |
| 8 | 1, 2, 3, 4, 12, 30, 31 |
| 9 | 8 |
| 10 | 79, 81, 83, 85, 90, 91, 92, 93, 94, 95, 98, 101, 103 |

^a The underlined substituents are those proposed in a previous study.¹⁰

be 0.92. The network was run for 30 epochs so as to ensure the stability of the results. It has to be noted that some substituents located at the borderline between two clusters have been found to fall in a neighbor cluster during separate runs, but this does not influence the conclusions that are drawn below from one of these runs. Indeed, for the sake of brevity the other runs are not detailed. They would simply provide complementary information.

4.2. Clustering Results. Membership of the 103 substituents to the ten clusters is given in Table 2. For comparison purposes, although the NLM was performed after a different scaling transformation, we have represented in Figure 4 the membership to the different clusters. It is noteworthy that despite the different preprocessing of the data,¹⁰ the results are generally in agreement. Indeed, the ART 2-A clusters allow one to logically divide the nonlinear map into 10 regions of physicochemically similar substituents. Some slight discrepancies with the NLM can be observed in clusters 5 and 10. Indeed, substituent 87 which belongs to cluster 5 is found in the middle of some substituents of cluster 10. This atypical assignment is due to its borderline location between the two clusters (see later). As regards substituents 79, 81, and 85, a reason for their

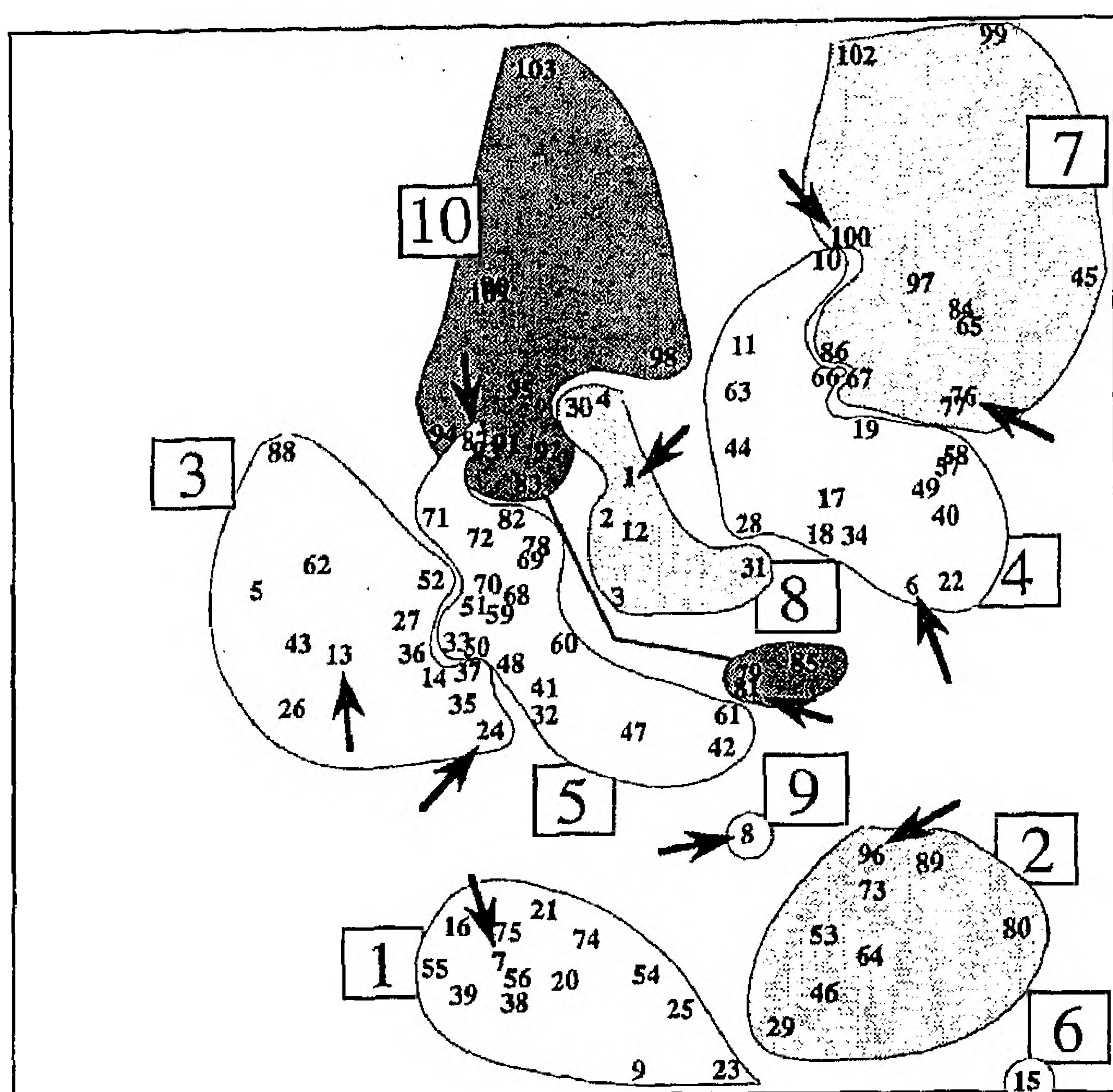


Figure 4. Representation on the nonlinear map of the 103 substituents¹⁰ of their cluster memberships derived by ART 2-A. See Table 1 for correspondence between the "small" numbers and the substituents. The large numbers inside a square correspond to the cluster indices given in Table 2. Cluster 10 is splitted into two groups linked by a line. The arrows indicate the substituents selected in a previous study.¹⁰

atypical position might be the different scaling procedures used in the two methods. Compared with NLM and HCA, the atypical character of substituents 8 and 15 is also stressed by the ART 2-A results since they are isolated. In the same way, comparison with the N2M (Figure 5) shows that the two methods provide comparable results. Indeed, in Figure 5, the presence in the N2M neurons (i.e., clusters) of the substituents contained in the clusters derived by ART 2-A are indicated in black. Thus, even if, the ART 2-A clusters can overlap two or more N2M neurons or if a same N2M neuron can contain substituents belonging to different ART 2-A clusters, the results of the two techniques are generally in accordance. Thus, for example, the substituents of the first ART 2-A cluster are all contained in neurons P and Q of the N2M display (Figure 5). Some discrepancies with the minimum spanning tree (MST) can, however, be observed for some ART 2-A clusters. Thus, for example, inspection of the map corresponding to cluster 3 shows that its substituents can be found in the neurons A and B which are directly linked by the MST but also in neuron E. However, it must be noted that in neuron E there is only substituent 88 from cluster 3 of ART 2-A. The inspection of its ρ^{calc} values (see next section for significance of ρ^{calc} values) shows that it is due to its borderline location between clusters 3, 5, and 10. This is confirmed on the nonlinear map (Figure 4). In the same way, for cluster 4, the substituents can be found in neurons I and J and also in neurons O and S. Figure 4 indicates that cluster 4 is between clusters 7 and 8 on the nonlinear map, and Figure 5 shows that substituents of cluster 7 are found in neurons M, O, and, S and that those of cluster 8 are in neuron I. Therefore, the fact that cluster 4 covers neurons common to both clusters 7 and 8 and neuron J is not surprising.

4.3. Interpretation of Numerical Results. To go further in the comparison, the key individuals of each cluster derived

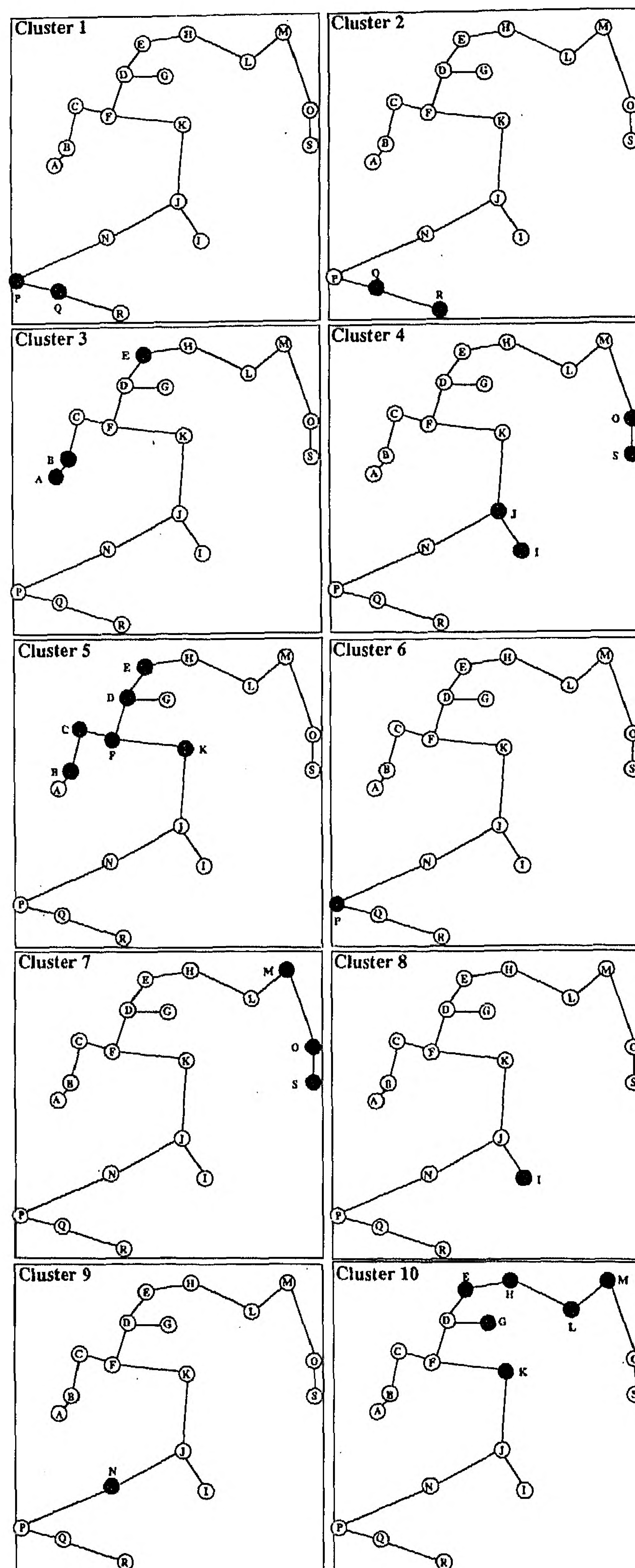


Figure 5. Representation on the N2M display¹² of the ART 2-A clusters. Each map corresponds to an ART 2-A cluster. For each map, the neurons in black are those containing the substituents included in the considered ART 2-A cluster. The cluster numbers correspond to those of Table 2.

by ART 2-A were determined using the ρ^{calc} values which depict the closeness of each individual to the cluster considered. They can also depict the similarity between any substituent and any cluster. For a given cluster, the substituent presenting the maximum ρ^{calc} value is the most representative of the cluster. The key individuals were compared to the selection proposed from HCA and NLM.¹⁰

Table 3. Comparison of Test Series Selected by HCA, NLM, and ART 2-A^a

| cluster no. | HCA, NLM ¹⁰ | ART 2-A |
|-------------|--|--|
| 1 | 7: OH | 7: OH (0.998) |
| 2 | 96: NHCOC ₆ H ₅ | 80: NHC ₄ H ₉ (0.998), 96: NHCOC ₆ H ₅ (0.960) |
| 3 | 13: CN, 24: OCH ₃ | 36: OCOCH ₃ (0.999), 14: SCN (0.998), 35: COCH ₃ (0.998), 37: CO ₂ CH ₃ (0.998), 13: CN (0.992), 24: OCH ₃ (0.992) |
| 4 | 6: H | 17: CH ₂ Br (0.999), 34: CH=CH ₂ (0.996), 6: H (0.952) |
| 5 | 87: OC ₆ H ₅ | 59: OCH(CH ₃) ₂ (0.999), 60: OC ₃ H ₇ (0.999), 87: OC ₆ H ₅ (0.992) |
| 6 | | 15: CO ₂ ⁻ (1.000) |
| 7 | 76: C ₄ H ₉ , 100: CH=CHC ₆ H ₅ -(trans) | 65: Si(CH ₃) ₃ (0.997), 84: C ₅ H ₁₁ (0.997), 76: C ₄ H ₉ (0.990), 100: CH=CHC ₆ H ₅ -(trans) (0.992) |
| 8 | 1: Br | 1: Br (0.998), 2: Cl (0.998), 12: CF ₃ (0.998), 30: CF ₂ CF ₃ (0.998) |
| 9 | 8: SH | 8: SH (1.00) |
| 10 | 81: N(C ₂ H ₅) ₂ | 95: OCOCH ₃ (0.999), 81: N(C ₂ H ₅) ₂ (0.992) |

^a The ρ^{calc} values of the substituents which are closest to the neural weight vector are given between brackets. The ρ^{calc} values of the substituents selected by HCA and NLM are also provided.

Table 4. Cosine of the Interclass Angles^a

| cluster no. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 1.000 | 0.918 | 0.743 | 0.354 | 0.735 | 0.685 | 0.281 | 0.402 | 0.755 | 0.686 |
| 2 | 0.918 | 1.000 | 0.636 | 0.336 | 0.683 | 0.663 | 0.382 | 0.226 | 0.664 | 0.708 |
| 3 | 0.743 | 0.636 | 1.000 | 0.574 | 0.989 | 0.838 | 0.488 | 0.639 | 0.396 | 0.943 |
| 4 | 0.354 | 0.336 | 0.574 | 1.000 | 0.611 | 0.092 | 0.954 | 0.915 | 0.577 | 0.661 |
| 5 | 0.735 | 0.683 | 0.989 | 0.611 | 1.000 | 0.839 | 0.563 | 0.616 | 0.387 | 0.980 |
| 6 | 0.685 | 0.663 | 0.838 | 0.092 | 0.839 | 1.000 | 0.078 | 0.118 | 0.077 | 0.790 |
| 7 | 0.281 | 0.382 | 0.488 | 0.954 | 0.563 | 0.078 | 1.000 | 0.764 | 0.495 | 0.666 |
| 8 | 0.402 | 0.226 | 0.639 | 0.915 | 0.616 | 0.118 | 0.764 | 1.000 | 0.616 | 0.592 |
| 9 | 0.755 | 0.664 | 0.396 | 0.577 | 0.387 | 0.077 | 0.495 | 0.616 | 1.000 | 0.380 |
| 10 | 0.686 | 0.708 | 0.943 | 0.661 | 0.980 | 0.790 | 0.666 | 0.592 | 0.380 | 1.000 |

^a 1 means identity and 0 means maximum dissimilarity.

The results are summarized in Tables 2 and 3. Inspection of Table 3 shows that the numerical values provided by ART 2-A and which guide the selection of the key individuals among the ten clusters allow one to find results similar to those obtained from HCA and NLM. Indeed, in Table 3, it is noteworthy that in all the clusters derived except the sixth one, we find at least one of the previously selected substituents. As regards cluster 6 which contains substituent 15 (i.e., CO₂⁻), it has to be considered that HCA also isolated this substituent in a cluster and that it was decided not to select it due to its atypical character.¹⁰ Besides this agreement between the different methods, one advantage of ART 2-A is that the preliminary selection of key substituents among each cluster is directed by an easily interpretable statistical parameter called ρ^{calc} (Table 3). Indeed, the largest values of ρ^{calc} correspond to the substituents that are the most characteristic of the cluster considered. If ever for any reason, such as difficulty for synthesis, a substituent was found not to be adequate, then the ρ^{calc} values can suggest the selection of another closest substituent since each cluster includes substituents with similar physicochemical profiles. Thus, for example, if in cluster 5, OCH(CH₃)₂ (substituent 59) was not satisfying, ART 2-A suggests to try with OC₃H₇ (substituent 60) which has the same ρ^{calc} value and possesses relatively similar physicochemical properties. Thus, the system is open, and freedom is left to the chemist to use any criterion for the selection of an optimal test series. The ρ^{calc} values can also be used to determine substituents that are located at the borderline between two clusters and, therefore, determine the relative position of each substituent. Thus, for example, it is noteworthy in Figure 4, that substituent 87 is placed in cluster 5, while it is among substituents belonging to cluster 10. Inspection of the ρ^{calc} (data not given) values shows that substituent 87 is actually

at the borderline between clusters 5 and 10 since its ρ^{calc} values for these two clusters are high and close to each other (i.e., $\rho^{\text{calc}} = 0.992$ for cluster 5 and $\rho^{\text{calc}} = 0.989$ for cluster 10). With regards to substituents 79, 81, and 85, the ρ^{calc} values also suggest a location at the borderline between clusters 5 and 10.

Other useful parameters for the interpretation of ART 2-A results, and the exploration of a data matrix are the interclass angles (Table 4). They provide information on the relationships between the clusters. Thus, the larger the cosine of the angles, the more similar the clusters. From a practical point of view, this allows us to select a substituent in another cluster if ever no appropriate substituents could be found in a given cluster. From Table 4, for example, if one wants to replace a substituent of cluster 1 by a substituent in another cluster, the interclass angles suggest cluster 2. This is verified on the nonlinear map (Figure 4).

4.4. Graphical Interpretation of Neural Weights. If the derivation of clusters and the selection of key substituents is very useful, it is also always necessary to have information on the clusters formed. This is achieved by inspection of the weights of each cluster providing an insight into the composition of the clusters in terms of physicochemical properties. They give physicochemical profiles of the substituents constituting the clusters. To summarize the information contained in the weight values and also to compare each cluster with all the others, an NLM of the ART 2-A clusters was performed from the weight values (Figure 6.1). Figure 6.1 visualizes the similarities between the different clusters with respect to their weights and, therefore, their physicochemical characteristics. Thus, for example, it clearly shows that clusters 3, 5, and 10 are close to each other compared with the other clusters. If graphical tools are used, one can in addition plot the weight data and,

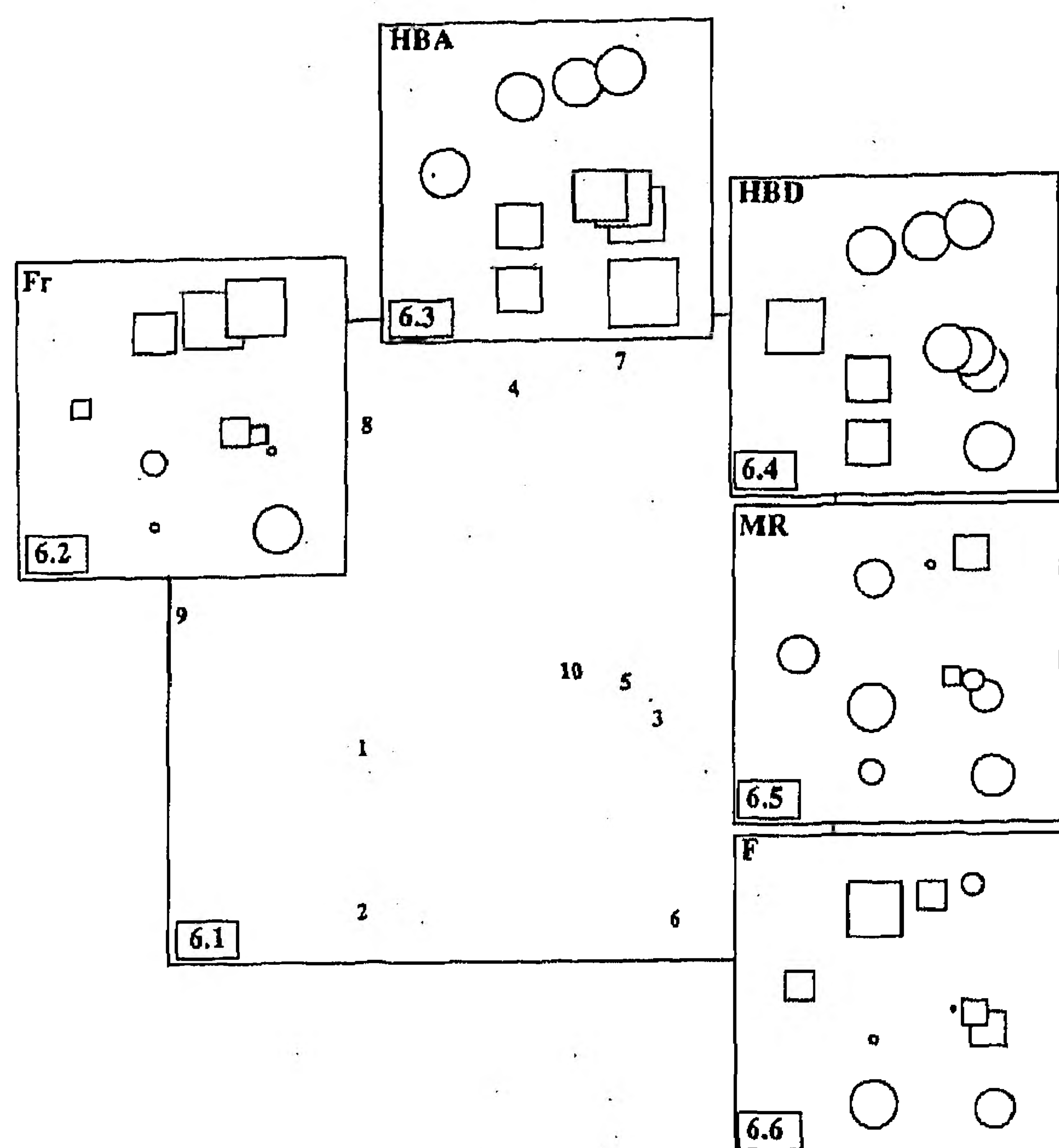


Figure 6. 6.1: NLM of ART 2-A weight values. The numbers correspond to the cluster indices given in Table 2 6.2–6.6: Plot of the weight values (after subtraction of the overall mean) linked to each of the five descriptors for each cluster of the nonlinear map (Figure 6.1). The larger the square, the larger the weight value, and the larger the circle, the lower the weight value.

therefore, visualize on a sole series of maps all the information contained in the original weight data table. Thus, in Figures 6.2–6.6, the weight data were plotted (after subtraction of the overall mean to improve the visualization) by means of squares (positive values) and circles (negative values) proportional in size to the transformed weights. In these figures, the larger the square, the larger the weight value, and the larger the circle, the lower the weight value. Although the complete analysis of these figures is beyond the scope of this paper, it can be noted, for example, that in Figure 6.2 substituents of the top clusters generally present high Fr values. In Figure 6.3, 6.4 obvious separations of the clusters, linked to HBA and HBD abilities, can be observed.

In addition, the nonlinear mapping of the clusters is an open approach, and any other useful information (e.g., biological or chemical information on the substituents contained in the clusters) for the interpretation or exploitation of the clusters in terms of structure–activity or structure–property relationships can be plotted.

5. CONCLUSIONS

The present study evidences the heuristic potency of ART 2-A for optimal test series design in QSAR. This technique which performs a nonlinear cluster analysis of data sets converges within a few epochs. In addition, ART 2-A offers the possibility of interpretation of the clusters derived in terms of the original variables by means of their weights which provide the characteristic profiles of the individuals constituting the clusters. Graphical tools are useful for visualizing the information condensed in a trained ART neural network. An interesting feature of these networks is also that by setting the vigilance parameter ρ^{\max} at different levels, it is possible to gradually inspect the structure of the data sets at hand. This is also made possible by the fact

that convergence is rapidly obtained. From a practical point of view, by simply selecting one or more substituents in each cluster, one obtains a highly informative test series. The ρ^{calc} values guide and facilitate the preliminary selection of substituents and also any deviation from this selection. ART 2-A results are consistent with those obtained from other techniques such as HCA, NLM, or N2M. All these methods have shown the ability to provide complementary and supplementary information. Last, although it was not shown in this study, ART 2-A can be used for predictive purposes. Thus, an unknown pattern can be presented to the network and its membership to a cluster determined. If its similarity with the existing clusters was not sufficient, another cluster would be formed and its relations with the others given by the different numerical data provided by the network.

REFERENCES AND NOTES

- (1) Domine, D.; Devillers, J.; Chastrette, M. A nonlinear map of substituent constants for selecting test series and deriving structure–activity relationships. 1. Aromatic series. *J. Med. Chem.* **1994**, *37*, 973–980.
- (2) Pleiss, M. A.; Unger, S. H. In *Comprehensive Medicinal Chemistry*, Vol. 4; Ramsden, C. A., Ed.; Pergamon Press: Oxford, 1990; pp 561–587.
- (3) Hansch, C.; Unger, S. H.; Forsythe, A. B. Strategy in drug design. Cluster analysis as an aid in the selection of substituents. *J. Med. Chem.* **1973**, *16*, 1217–1222.
- (4) Dove, S.; Streich, W. J.; Franke, R. On the rational selection of test series. 2. Two-dimensional mapping of intraclass correlation matrices. *J. Med. Chem.* **1980**, *23*, 1456–1459.
- (5) Alunni, S.; Clementi, S.; Edlund, U.; Johnels, D.; Hellberg, S.; Sjöström, M.; Wold, S. Multivariate data analysis of substituent descriptors. *Acta Chem. Scand.* **1983**, *B37*, 47–53.
- (6) van de Waterbeemd, H.; El Tayar, N.; Carrupt, P. A.; Testa, B. Pattern recognition study of QSAR substituent descriptors. *J. Comput.-Aided Mol. Design* **1989**, *3*, 111–132.
- (7) Tosato, M. L.; Geladi, P. In *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, 1990; pp 317–341.
- (8) Baroni, M.; Clementi, S.; Cruciani, G.; Kettaneh-Wold, N.; Wold, S. D-optimal designs in QSAR. *Quant. Struct.-Act. Relat.* **1993**, *12*, 225–231.
- (9) Caliendo, G.; Greco, G.; Novellino, E.; Perissutti, E.; Santagada, V. Combined use of factorial design and comparative molecular field analysis (CoMFA): A case study. *Quant. Struct.-Act. Relat.* **1994**, *13*, 249–261.
- (10) Domine, D.; Devillers, J.; Chastrette, M. A nonlinear map of substituent constants for selecting test series and deriving structure–activity relationships. 2. Aliphatic series. *J. Med. Chem.* **1994**, *37*, 981–987.
- (11) Devillers, J. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; pp 255–263.
- (12) Domine, D.; Wienke, D.; Devillers, J.; Buydens, L. In *Neural Networks in QSAR and Drug Design*; Devillers, J., Ed.; Academic Press: London, 1996; pp 223–253.
- (13) Putavy, C.; Devillers, J.; Domine, D. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: London, 1996; pp 243–269.
- (14) Wienke, D. In *Chemometrics: Exploring and Exploiting Chemical Information*; Buydens, L., Melssen, W., Eds.; University Press: University of Nijmegen, 1994; pp 197–222.
- (15) Wienke, D.; Kateman, G. Adaptive resonance theory based artificial neural networks for treatment of open-category problems in chemical pattern recognition—Application to UV/VIS- and IR-spectroscopy. *Chemom. Intell. Lab. Syst.* **1994**, *23*, 309–329.
- (16) Wienke, D.; Xie, Y.; Hopke, P. K. An adaptive resonance theory based artificial neural network (ART-2a) for rapid classification of airborne particles by their scanning electron microscopy images. *Chemom. Intell. Lab. Syst.* **1994**, *26*, 367–387.
- (17) Wienke, D.; van den Broek, W.; Feldhoff, R.; Huth-Fehre, T.; Kantimm, T.; Quick, L.; Melssen, W.; Winter, F.; Cammann, K.; Buydens, L. Adaptive resonance theory based neural network for supervised chemical pattern recognition (FuzzyARTMAP). Part 2: Classification of post-consumer plastics by remote NIR spectroscopy. *Chemom. Intell. Lab. Syst.* **1996**, (in press).
- (18) Wienke, D.; van den Broek, W.; Melssen, W.; Buydens, L.; Feldhoff, R.; Huth-Fehre, T.; Kantimm, T.; Quick, L.; Winter, F.; Cammann, K. Comparison of an adaptive resonance theory based artificial neural network with other classifiers for fast sorting of post-consumer plastics

- by remote NIR sensing with an InGaAs diode detector array. *Anal. Chim. Acta* **1995**, 317, 1–16.
- (19) Xie, Y.; Hopke, P. K.; Wienke, D. Airbone particle classification with a combination of chemical composition and shape index utilizing an adaptive resonance artificial neural network. *Environ. Sci. Technol.* **1994**, 28, 1921–1928.
- (20) Wienke, D.; Buydens, L. An adaptive resonance theory based artificial neural network for supervised chemical pattern recognition (Fuzzy-ARTMAP). Part 1: Theory and basic properties. *Chemom. Intell. Lab. Syst.* **1996**, (in press).
- (21) Wienke, D.; Buydens, L. Adaptive resonance theory neural networks—The “ART” of real-time pattern recognition in chemical process monitoring? *Trends Anal. Chem.* **1995**, 99, 1–8.
- (22) Wienke, D.; Domine, D.; Buydens, L.; Devillers, J. In *Neural Networks in QSAR and Drug Design*; Devillers, J., Ed.; Academic Press: London, 1996; pp 119–156.
- (23) Grossberg, S. Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biol. Cybern.* **1976**, 23, 121–134.
- (24) Grossberg, S. Adaptive pattern classification and universal recoding, II: Feed-back, expectation, olfaction, and illusions. *Biol. Cybern.* **1976**, 23, 187–203.
- (25) Grossberg, S. *Studies of Mind and Brain*; D. Reidel Publishing Company: Dordrecht, 1982.
- (26) Carpenter, G. A.; Grossberg, S.; Reynolds, J. H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* **1991**, 4, 565–588.
- (27) Carpenter, G. A.; Grossberg, S. ART 2: Self-organization of stable category recognition codes for analog input patterns. *Appl. Optics* **1987**, 26, 4919–4930.
- (28) Carpenter, G. A.; Grossberg, S.; Rosen, D. B. ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks* **1991**, 4, 493–504.
- (29) Carpenter, G. A.; Grossberg, S. ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks* **1990**, 3, 129–152.
- (30) Carpenter, G. A.; Grossberg, S.; Rosen, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks* **1991**, 4, 759–771.
- (31) Carpenter, G. A.; Grossberg, S.; Markuzon, N.; Reynolds, J. H.; Rosen, D. B. FuzzyARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. Neural Networks* **1992**, 3, 698–713.
- (32) Carpenter, G. A.; Grossberg, S. *Pattern Recognition by Self-Organizing Neural Networks*; The MIT Press: Cambridge, Massachusetts, 1991; p 691.
- (33) Kohonen, T. *Self-Organization and Associated Memory*; Springer-Verlag: Heidelberg, 1989.
- (34) Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, 78, 1464–1480.
- (35) Alpsan, D. In *Proceedings of IEEE International Conference on Neural Networks. June 28–July 2, 1994, Volume 6*; IEEE: Piscataway, 1994; pp 3512–3515.
- (36) Kane, J. S.; Paquin, M. J. POPART: Partial optical implementation of adaptive resonance theory 2. *IEEE Trans. Neural Networks* **1993**, 4, 695–702.
- (37) Caglayan, A. K.; Snorrason, M. In *World Congress on Neural Networks. Oregon Convention Center, Portland, Oregon, July 11–15, 1993*; Lawrence Erlbaum Associates: Hillsdale, 1993; pp II-532–II-535.
- (38) Gonsalves, P. G.; Snorrason, M.; Caglayan, A. K. In *World Congress on Neural Networks. Oregon Convention Center, Portland, Oregon, July 11–15, 1993*; Lawrence Erlbaum Associates: Hillsdale, 1993; pp II-419–II-422.
- (39) Snorrason, M.; Caglayan, A. K. In *World Congress on Neural Networks. Town & Country Hotel, San Diego, California, USA, June 5–9, 1994*; Lawrence Erlbaum Associates: Hillsdale, 1994; pp IV-240–IV-245.
- (40) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.
- (41) Domine, D.; Devillers, J.; Chastrette, M.; Karcher, W. Non-linear mapping for structure–activity and structure–property modelling. *J. Chemometrics* **1993**, 7, 227–242.

CI960376P